Amendments to the Specification:

Please replace the paragraph beginning on page 1, line 13 with the following rewritten paragraph:

a1

-- In many environments, traditional hands-on user interfaces, for example, a mouse and keyboard, for interacting with a computer are not practical. One example of such an environment is an operating theater (OT) where there is a need for strict sterility. A surgeon, and everything coming into contact with his/her hands must be sterile. Therefore, the mouse and keyboard ~~maybe~~ may be excluded from consideration as an interface because they may not be sterilized. --

Please replace the paragraph beginning on page 2, line 3 with the following rewritten paragraph:

a2

-- Areas of human-machine interaction in the OT include, for example, voice recognition and gesture recognition. There are several ~~commercially~~ commercial voice recognition systems available. In the context of the OT, their advantage is that the surgeon can continue an activity, for example, a suture, while commanding the imaging system. However, the disadvantage is that the surgeon needs to mentally translate geometric information into language: e.g., “turn right”, “zoom in”, “stop”. These commands need to include some type of qualitative information. Therefore, it can be complicated and tiresome to achieve a specific 3D orientation. Other problems related to voice recognition are that it may fail in a noisy environment, and the system may need to be trained to each user. --

Please replace the paragraph beginning on page 3, line 4 with the following rewritten paragraph:

a3

--For feature detection/extraction, applications may use color to detect human skin. An advantage of a color-based technique is real-time performance. However, the variability of skin color in varying lighting conditions can lead to false detection. Some applications use motion to localize the gesture. A drawback of a motion cue approach is that assumptions ~~maybe~~ may be needed to make the system operable, e.g., a stationary background and one active gesturer. Other methods, such as using data-gloves/sensors to collect 3D data, may not be suitable for a human-machine interface because they are not natural. –

Please replace the paragraph beginning on page 5, line 13 with the following rewritten paragraph:

a4

-- Determining the trajectory includes determining a difference in a size of the object over a pre-determined time period, determining a plurality of angles between a plurality of lines connecting successive ~~controids~~ centroids over the time period, and determining a feature vector according to the angles and lines.[.]--

Please replace the paragraph beginning on page 10, line 10 with the following rewritten paragraph:

a5

-- Consider two valid gesture commands, move left and move right. Both commands may need the user's hand to be waved horizontally and the user can continue this movement as many times as desired. Given no information about where the movement starts, there maybe no way to distinguish between the motion trajectory patterns, e.g., left or right waves. Similar ambiguities can occur when other translations are performed. For this reason, the system and method needs to know or determine a starting point for a gesture command. According to an embodiment of the present invention, by holding the hand stationary before performing a new gesture, the stationary point becomes a reference point. The reference point is used to distinguish among, for example, moving left or right, up or down, and forward or backward.--

Please replace the paragraph beginning on page 11, line 12 with the following rewritten paragraph:

a6

-- The system and method can detect changes in a background by determining an intensity of each image from video stream. To eliminate noise, a Gaussian filter can be applied to each image. A gradient map of pixel intensity can be determined. After determining the gradient map of a current image frame, the gradient ~~may~~ map is compared with the learned background gradient map. If a given pixel differs less than a threshold between these two gradient maps, the pixel is determined to be a background pixel, and can be marked accordingly. A pre-determined threshold can be used. One with ordinary skill in the art would appreciate, in light of the present invention, that additional methods for selecting the threshold exist, for example, through knowledge of sensor characteristics or through normal illumination changes allowed in the background. According to an embodiment of the present invention the largest area of connected background pixels can be treated as background region.--

Please replace the paragraph beginning on page 13, line 10 with the following rewritten paragraph:

a7

-- The trajectory of the centroid of the detected skin object is often used as the motion trajectory of the object. However, it has been determined that there are many objects having skin-like color in an office environment. For example, a wooden bookshelf or a poster on a wall may be misclassified as a skin-like object. Therefore, the system and method attempts to eliminate background pixels as discussed in above. Besides, the skin objects (user's hand and probably the arm) are sometimes split up into two or more blobs. Other skin regions such as face may also appear in the view of the camera. These problems together with non-uniform illumination make the centroid vary dramatically and ~~leads~~ lead to false detections. For these reasons, a stable motion trajectory is hard to obtain by just finding the largest skin area. To handle these problems, a temporal likelihood can be defined as $L^t(x, y, t)$ of each pixel $I(x, y)$ as:

a7 cont'd

$$L^t(x, y, t) = \lambda L(x, y) + (1 - \lambda)\ L^t(x, y, t\text{-}1) \qquad (3)$$

where λ is a decay factor. Experiments show that a value of λ equal to 0.5 can be used.--

Please replace the paragraph beginning on page 14, line 20 with the following rewritten paragraph:

a8 C'

-- Recognition of a user's hand motion patterns can be accomplished using TDNN according to an embodiment of the present invention. Experiments show that TDNN has good performance on motion pattern classification. As shown by experiments, TDNN has better performance if the number of output labels was kept small. Another advantage is that a small number of output labels makes networks simple and saves time at network training stage. For these reasons user's gestures are tested hierarchically. Further, TDNN applied hierarchically[,] has been determined to be suitable for the classification of the eight motion patterns described above. For instance, left movement and right movement have the common motion pattern of horizontal hand movement. Thus, once horizontal movement is detected, the range of the motion is compared with the reference point to differentiate these two gestures.--

Please replace the paragraph beginning on page 15, line 20 with the following rewritten paragraph:

a9

-- Suppose the length of an input pattern is w, the feature vectors $\{v_{t-w+1}, v_{y-w+2},\ldots, v_t\}$ from $\{c_{t-w}, c_{t-w+1},\ldots,c_t\}$ are extracted to form a TDNN input pattern. When the maximum response from the network is relatively small, as compared with other label responses, the input pattern is classified as an unknown. Some false detections or unknowns are inevitable. False detection can occur when the trajectory of a translation ~~are~~ is similar to an arc of a circle. To minimize false detection and obtain stable performance, a fixed number of past results are checked. When more than half of these past results indicate the same output pattern, this output pattern is determined to be a final result. This method has been used to successfully obtain a reliable recognition rate.--

Please replace the paragraph beginning on page 17, line 7 with the following rewritten paragraph:

a10

-- For the detection of circular movements, the angle between vector c_tc_{t-1} and vector $c_{t-1}c_{t-2}$ is computed as the feature vector 406. This feature can distinguish between clockwise and counterclockwise circular movements. As expected, users can draw circles from any position. In particular, a spiral would be classified as one of the circular movements instead of a translation. Referring to Fig. 4, the method can use a voting method 407 to check past results to form meaningful output, the system decreases the possibility of false classification. The method determines whether a given gesture is a valid gesture command 408. A valid ~~gestures~~ gesture needs to be performed continually in some time interval to initialize the command.--

Please replace the paragraph beginning on page 17, line 20 with the following rewritten paragraph:

a11

-- Figs. 5 and 6 show some examples of our experimental results. In each image, the black region, e.g., 501, is viewed as background. The bounding box, e.g., 502 (highlighted in white in Fig. 5b for clarity), of each image indicates the largest skin area as determined by thresholded likelihood, Equation (4) ~~(2)~~. Note that bounding boxes are only used for display. The arrow(s), e.g., 503, on each bounding box show the classification result. A bounding box with no arrow, for example, as in Figs. 5a-c, on it means that the gesture is an unknown pattern, or that no movement has occurred, or insufficient data has been collected. Because we classify motion patterns along windows in time, there may be some delay after a gesture is initialized (data is not sufficient for system to make a global decision). --